Journal of Water and Soil
Vol. 35, No. 5, Dec.-Jan. 2021, p. 719-734

نشریه آب و خاک (علوم و صنایع کشاورزی)
جلد ۳۵، شماره ۵، آذر – دی ۱۴۰۰، ص. ۷۳۴-۷۱۹

مقاله پژوهشی

مدل‌سازی تغییرات عمقی کربنات کلسیم معادل خاک با استفاده از الگوریتم‌های یادگیری ماشین در دشت قزوین

سید روح اله موسوی[1]- فریدون سرمدیان[2]*- محمود امید[3]- پاتریک بوگارت[4]

تاریخ دریافت: ۱۴۰۰/۰۵/۲۵
تاریخ پذیرش: ۱۴۰۰/۰۷/۰۳

چکیده

کربنات کلسیم معادل یکی از ویژگی‌های کلیدی خاک‌های مناطق خشک و نیمه‌خشک است که بررسی تغییرات سطحی و عمقی آن از اهمیت ویژه‌ای در بهره‌برداری پایدار از خاک‌های زراعی برخوردار است. هدف از این تحقیق مدل‌سازی مکانی کربنات کلسیم معادل (CCE) در پنج عمق استاندارد ۱۰۰-۶۰، ۶۰-۳۰، ۳۰-۱۵، ۱۵-۵ و ۵-۰ سانتی‌متر متناظر با پروژه جهانی نقشه خاک با استفاده از سه الگوریتم یادگیری ماشین جنگل تصادفی (RF)، رگرسیون درخت تصمیم (DTr) و k-نزدیک‌ترین همسایگی (k-NN) بود. مطالعات میدانی و آزمایشگاهی شامل حفر ۲۷۸ خاکرخ، نمونه‌برداری و انجام تجزیه‌های فیزیکوشیمیایی موردنظر بود. متغیرهای کمکی شامل مشتقات مدل رقومی ارتفاع، شاخص‌های سنجش‌ازدور، داده‌های اقلیمی و خاک بودند که انتخاب دسته مناسب آن‌ها با استفاده از روش تجزیه مؤلفه‌های اصلی (PCA) و نظر کارشناس انجام گردید. همسان‌سازی مقادیر CCE در اعماق استاندارد به‌وسیله تابع عمق اسپیلاین اجرا گردید. بر اساس روش PCA در مؤلفه‌های اول تا پنجم با توجیه بیش از ۸۰٪ واریانس تجمعی، متغیرهای کمکی شاخص هموار‌ی دره با وضوح مکانی بالا (MrVBF)، میانگین دمای سالانه (MAT)، شاخص سبزینگی (Greenness)، احتمال افق کلسیک (Cal.hr) و شاخص اثر باد (Wind Effect) و براساس نظر کارشناس، درصد رس (Clay) انتخاب گردیدند. الگوریتم RF در مقایسه با دو الگوریتم دیگر (k-NN،DTr) با دامنه مقادیر R^2 برابر ۰/۸۳ – ۰/۷۶ و RMSE برابر ۲/۱۴- ۳/۲۱ درصد بالاترین میزان دقت و حداقل خطا را ارائه نمود. در سه عمق سطحی تغییرات مکانی CCE متأثر از متغیر Clay بود، در حالی‌که در اعماق زیرین Cal.hr مهم‌ترین فاکتور پیش‌بینی کننده آن بود. به‌طورکلی استفاده از رویکردهای نوین نقشه‌برداری در تهیه نقشه CCE به دلیل تأثیر این ویژگی بر روی قابلیت دسترسی رطوبت خاک و جذب عناصر غذایی توسط گیاهان بسیار کاربردی است.

واژه‌های کلیدی: اعماق استاندارد، تابع اسپیلاین، فاکتورهای خاک‌سازی، نقشه‌برداری رقومی خاک

مقدمه

نقشه‌های خاک بیانگر نمایش ساده‌ای از الگوهای پیچیده و ناشناخته پراکنش مکانی ویژگی‌های خاک است (۲۸). استفاده از روش‌های نقشه‌برداری رقومی خاک باهدف تولید نقشه‌های با ماهیت پیوسته به همراه اعتبارسنجی کمی و محاسبه عدم قطعیت آن‌ها یکی از روش‌های کاهش هزینه و زمان مطالعات خاک‌شناسی است (۶).

در این رویکرد از الگوریتم‌های یادگیری ماشین برای شناسایی رابطه بین داده‌های خاک و متغیرهای کمکی استفاده می‌شود و انتخاب بهینه الگوریتم پیش‌بینی کننده مکانی تأثیر بالایی در صحت نهایی نقشه‌های تولیدشده دارد به طوری‌که در یک بررسی جامع از مطالعات ۱۰ سال اخیر توسط خالدیان و همکارن (۱۳) الگوریتم‌های رگرسیون خطی چندمتغیره (MLR)، K - نزدیک‌ترین همسایگی (k-NN)، رگرسیون ماشین بردار پشتیبان (SVR)، کوبیست (Cubist)، جنگل تصادفی (RF) و شبکه‌های عصبی مصنوعی (ANN) بر اساس پنج فاکتور،کمی‌سازی فرا پارامترها، اندازه نمونه،

۱ و ۲- به‌ترتیب دانشجوی دکتری مدیریت منابع خاک و استاد، گروه علوم و مهندسی خاک، پردیس کشاورزی و منابع طبیعی، دانشگاه تهران، کرج، ایران
(*- نویسنده مسئول: Email: fsarmad@ut.ac.ir)
۳- استاد گروه مهندسی ماشین‌های کشاورزی، پردیس کشاورزی و منابع طبیعی، دانشگاه تهران، کرج، ایران
۴- استاد دانشکده محیط زیست و علوم زمین، دانشگاه کاتولیک لوون، دی لوون، بلژیک

DOI: 10.22067/JSW.2021.71748.1076

انتخـاب متغیـر، زمـان یـادگیری و تفسـیرپذیری ارزیـابی نمـوده و درمجموع الگوریتم RF را نسبت دیگر الگوریتم‌ها به‌عنوان مدل برتر معرفی نمودند. تاکنون استفاده از الگوریتم‌های یـادگیری ماشـین در مدل‌سازی پیش‌بینی مکانی ویژگی‌هـای خـاک در اعمـاق سـطحی مورد توجه پژوهشگران قرارگرفته است (۴۵، ۳۷، ۲۹، ۲۷، ۲۱ و ۱۷)، و باوجود تغییرات پیوسته ویژگی‌ها با عمق، مدل‌سـازی آن‌هـا کمتـر موردتوجه واقع‌شده است (۹، ۱۸ و ۳۳).

مدل‌سازی تغییرات عمقی خصوصیات خاک بـا توجـه بـه تغییـر ضخامت افق‌های خاک در گستره وسیع نیـاز بـه اسـتفاده از توابـع هماهنگ‌سازی را توجیه نموده است که از مهم‌ترین آن‌ها مـی‌تـوان به توابع اسپیلاین با سطوح یکسان اشاره نمـود (۳۹، ۳۸، ۱۸ و ۲۳). کربنات کلسیم معادل (CCE) یکی از ویژگی‌های مهـم خـاک‌هـای مناطق خشک و نیمه‌خشک است که مطالعات انـدکی در ارتبـاط بـا تغییرپذیری مکانی سـطحی و عمقـی آن بـه‌عنـوان یکـی از منـابع پتانسیل میزان کربن غیر آلی ذخیره‌شده در خاک انجـام شـده اسـت (۳۵ و ۱). در پژوهشی، تغییرات متوسط وزنـی CCE تـا عمـق ۱۰۰ سانتی‌متر در دشت شهرکرد با اسـتفاده از مـدل‌هـای SVR، RF و ANN و متغیرهای محیطی مساحت حـوزه آبخیـز و سـطح اسـاس شبکه کانال[1] بررسی و ANN به‌عنوان مدل با خطای کمتـر معرفـی گردیـد (۶). همچنـین اسـرویانس و همکـاران (۳۵) از مـدل RF در مقیاس قاره هند برای مدل‌سازی چگـالی کربنـات کلسـیم تـا ۱۰۰ سانتی‌متر بدون لحاظ توابع عمقی استفاده نمـوده و دقـت بـالایی را گزارش نمودند. در مطالعات ایران نیز زراعت پیشه و همکاران (۴۵) و مصلح و همکاران (۲۰) در دشت‌های فلات مرکزی از مدل‌های RF، Cubist، MLR برای مدل‌سـازی CCE سـطحی اسـتفاده نمـوده و دقت متوسط به پایین را گزارش نمودند. همچنـین در منطقـه مـورد مطالعه (۲۲ و ۷) با استفاده از روش‌های زمین‌آماری اقـدام بـه تهیـه نقشه تغییرات مکانی CCE در خاک سطحی نموده‌اند و نتایج خـوب تا متوسطی را گـزارش نمـوده‌انـد. امیریـان چکـان و همکـاران (۱) تغییرات عمودی و جانبی کربنات کلسیم معادل در دشت سیلاخور بـا استفاده از توابع اسپیلاین و زمین‌آمار بررسی کردند و نتایج حـاکی از دقت خوب توابع اسپیلاین در تخمین عمقی CCE است.

با این‌وجود مطالعـات محـدودی در ارتبـاط بـا بررسـی تغییـرات عمقی و جانبی CCE با استفاده هم‌زمان از توابـع عمـق اسـپیلاین، مدل‌های یادگیری ماشین، متغیرهای محیطی و ویژگی‌هایی از جنس خود خاک به‌عنوان فاکتور "s" در معادله اسکورپن صـورت پذیرفتـه است؛ بنابراین پـژوهش حاضـر بـا اهـداف ۱) مقایسـه کـارایی سـه الگوریتم یادگیری ماشین RF، DTr و k-NN در تهیـه نقشـه‌هـای پیش‌بینی مکانی CCE در اعماق اسـتاندارد ۱۰۰–۶۰، ۶۰–۳۰، ۳۰–۱۵، ۱۵–۵ و ۵–۰ سانتی‌متـر و ۲) شناسـایی مهـم‌تـرین متغیرهـای کمکـی پـیش‌بینـی کننـده CCE در خـاک‌هـای منـاطق خشـک و نیمه‌خشک صورت پذیرفت.

1- Channel network base level

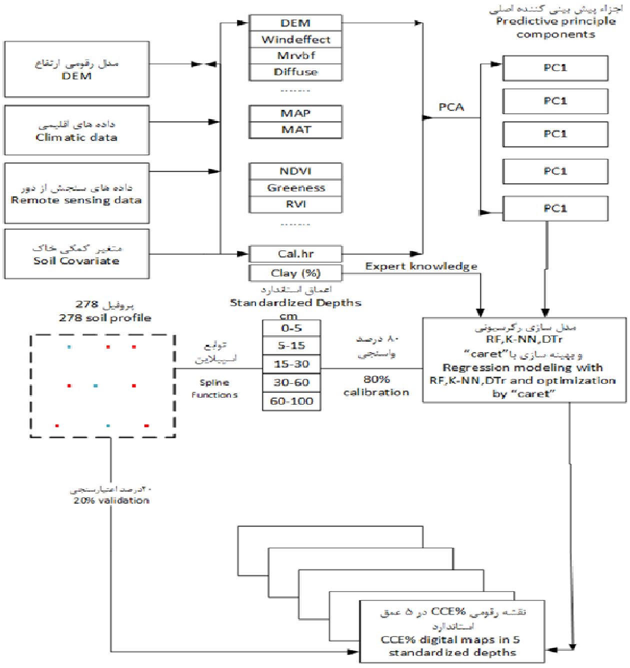
مواد و روش‌ها

پژوهش حاضر در بخشی از اراضـی دشـت قـزوین بـا مسـاحت حدود ۶۰۰۰۰ هکتار در حدفاصل دو اسـتان قـزوین و البـرز صـورت پذیرفت (شکل ۱). منطقه مورد مطالعـه در حدواسـط موقعیـت‌هـای جغرافیـایی″۳۵/۰۶ ′۱۵ °۵۰ و ″۲۵/۵۳ ′۲۹ °۵۰ طـول شـرقی و ″۵۴′۳۰/۳۸ °۳۵ تا ″۳۸/۸۳ ′۵۴ °۳۶ عرض شمالی واقع شده اسـت. بر اساس اطلاعات اقلیمی بلندمـدت ۴۹ سـاله (سـازمان هواشناسـی کشور) منطقه دارای میانگین بـارش ۲۸۰ میلـی‌متـر و دمـای ۱۴/۵ درجه سانتی‌گراد بوده کـه بـه ترتیـب دارای رژیـم‌هـای رطـوبتی و حرارتی زریک[2]، اردیک[3]، آکوئیک[4] و ترمیک[5] می‌باشـد. بـر اسـاس مطالعات میدانی و آنالیزهای آزمایشگاهی خاک‌های منطقه در هشت گروه بزرگ طبقه‌بندی می‌شوند که سه کـلاس هـاپلوزرپتز[6]، کلسـی زرپتز[7] و هاپلوکلسیدز[8] بیشترین مساحت منطقه مورد مطالعه را شامل می‌شوند (۳۶). کاربری غالب اراضی منطقه شامل زراعت آبی شـامل محصـولات یکسـاله و چندسـاله گنـدم، جـو، ذرت، یونجـه، کلـزا، صیفی‌جات و محصولات باغی درختان انگور، گردو، گلابـی و سـیب به‌همراه مراتع شور و غیر شور می‌باشند.

روندنمای کلی تحقیق

در راستای توصیف هر چه بهتر چارچوب مـورد اسـتفاده در ایـن تحقیق روندنمای[9] کلی آن در شکل ۱ ارائه شده است؛ که مهمترین گام‌های آن شامل ۱) تهیه متغیرهای محیطی با قابلیـت دسـت‌یـابی آسان و حداقل هزینه ۲) انتخاب مناسب‌ترین متغیرهای محیطـی بـا روش تجزیه مولفه‌های اصلی و نظر کارشناس ۳) مطالعات میدانی و نمونه‌برداری از خاکرخ‌های مشـاهداتی ۴) همسـان‌سـازی[10] مقـادیر CCE در افق‌های ژنتیکی بر اساس اعماق استاندارد ۵)

2- Xeric
3- Aridic
4- Aquic
5- Thermic
6- Haploxerepts
7- Calcixerepts
8- Haplocalcids
9- Flowchart
10- Harmonization



شکل ۱- روندنمای کلی تحقیق

Figure 1- The flowchart of this research

مدلسازی مکانی CCE با استفاده از الگوریتم‌های RF، DTr و k-NN ۶) اعتبارسنجی نتایج مدل‌سازی با استفاده از آماره‌های اعتبارسنجی ۷) تهیه نقشه‌های پراکنش مکانی CCE در اعماق استاندارد متناظر با پروژه جهانی نقشه خاک[1] بر اساس مناسب‌ترین الگوریتم یادگیری ماشین انجام گردید.

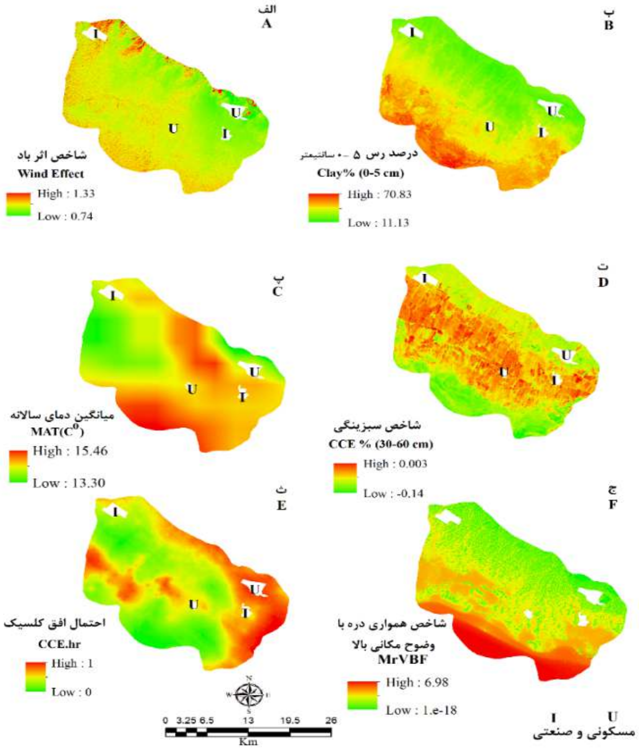
متغیرهای محیطی

در این تحقیق مجموعه‌ای از متغیرهای محیطی با قابلیت دسترسی آسان و هزینه حداقل که شامل مشتقات اولیه و ثانویه مدل رقومی ارتفاع با وضوح مکانی ۱۲/۵ متر نماینده فاکتور "r"، شاخص‌های قابل تهیه از نسبت‌گیری باندهای ماهواره لندست ۸ با وضوح مکانی ۳۰ متر نماینده فاکتور "o" و اجزا بافت خاک (درصد رس، سیلت و شن) و نقشه احتمال افق مشخصه کلسیک[2] (Cal.hr) به عنوان نماینده فاکتور "s" در مدل اسکورپن استفاده گردید. در مجموع ۲۲ متغیر محیطی از منابع اشاره شده تهیه گردیدند (جدول ۱) که شش مورد از مهمترین آنها در شکل ۲ ارائه شده است. برای انتخاب مناسبترین متغیرهای کمکی از یک رویکرد نیمه خودکار[3] – نظارت نشده، به‌نام تجزیه مولفه‌های اصلی[4] (PCA) استفاده گردید (۱۳). مولفه‌های اصلی با ارزش ویژه[5] بزرگتر از "یک" مورد بررسی قرار گرفتند. درون هر مولفه اصلی (PC) متغیرهای کمکی که دارای بالاترین مقادیر بودند به همراه سایر شاخص‌هایی که در ۱۰٪ با آنها باشند، شناسایی شدند. سپس بر اساس تطبیق ویژگی‌های منتخب در هر PC با ضریب همبستگی بین آنها، در صورت وجود همبستگی بالا (با فرض ضریب همبستگی بیشتر از ۰/۶) بین آنها در نهایت متغیر کمکی که بالاترین مقدار را داشت، نگه داشته شده و بقیه حذف گردیدند (۳۴). تجزیه مولفه‌های اصلی در نرم‌افزار SPSS نسخه ۲۱ انجام گردید.

مطالعات نمونه‌برداری خاک و تجزیه و تحلیل آزمایشگاهی

در این پژوهش نمونه‌برداری از مجموع ۲۷۸ خاکرخ انجام گرفت (شکل ۳).

1- Global soil map project
2- Calcic diagnostic horizon
3- Semi-automated
4- Principal component analysis
5- Eigenvalues



شکل ۲- نمونه‌ای از مهمترین متغیرهای محیطی

Figure 2- Example of the most importance environmental covariates

موقعیت نقاط نمونه‌برداری با توجه به وضعیت زمین‌نما (فیزیوگرافی، کاربری اراضی، سهولت دسترسی) بر اساس روش طبقه‌بندی تصادفی و نظر کارشناس طراحی و سپس با استفاده از سیستم موقعیت یاب جهانی[1] (مدل گارمین ۶۲ اس[2]) جانمایی و حفر گردیدند. نمونه‌برداری خاک از کلیه افق‌های ژنتیکی قابل شناسایی انجام و در ادامه برای انجام آزمایش‌های مورد نظر به آزمایشگاه منتقل گردیدند. پس از هوا خشک شدن نمونه‌های خاک و عبور آنها از الک ۲ میلی‌متری، مقادیر درصد کربنات کلسیم معادل (CCE) برای ۸۳۵ نمونه خاک جمع‌آوری شده به روش تیتراسیون (۲۵) اندازه‌گیری گردید.

استاندارسازی تغییرات عمق خاک

به‌منظور هماهنگ‌سازی نتایج مدل سازی در این تحقیق با مطالعات جهانی نقشه‌های خاک (۹)، تهیه و مقادیر CCE خاک‌ها بر اساس اصول تعریف شده توسط (۲) در پنج عمق ۶۰-۱۰۰، ۳۰-۶۰، ۱۵-۳۰، ۵-۱۵، ۰-۵ سانتی‌متر استاندارد گردیدند. به‌منظور استانداردسازی و همگن‌سازی، معادله عمق اسپیلاین[3] با سطح برابر به داده‌های CCE در هریک از افق‌های ژنتیکی برازش داده شد (شکل ۴). خلاصه آماری داده‌های CCE خاک در هر یک از اعماق استاندارد (جدول ۱) در نرم‌افزار تجزیه و تحلیل آماری R نسخه 4.0.3 انجام گردید.

مدل‌سازی مکانی

در این تحقیق از سه الگوریتم یادگیری ماشین شامل جنگل تصادفی (RF)، رگرسیون درخت تصمیم (DTr) و k-نزدیکترین همسایگی (k-NN) به منظور مدل‌سازی مکانی تغییرات عمقی CCE در پنج عمق استاندارد ۱۰۰-۶۰، ۶۰-۳۰، ۳۰-۱۵، ۱۵-۵ و ۵-۰ سانتی‌متر استفاده گردید. کلیه مراحل مدل‌سازی در محیط نرم‌افزار متن باز آماری R و با استفاده از بسته‌های تخصصی "caret"، "random Forest" و "rpart" صورت پذیرفت.

1- Geographic position system
2- Garmin 62s
3- Spline depth function

جدول ۱- متغیرهای محیطی مورد استفاده برای مدل‌سازی مکانی CCE%
Table 1- Environmental covariates applied for spatial modeling of CCE%

فاکتور خاک‌سازی Soil forming factors	متغیر کمکی محیطی Environmental covariates	نام متغیر و نماد Covariate and symbol
توپوگرافی Relief (r)	مدل رقومی ارتفاع	Digital elevation model
	شاخص اثر باد	Wind effect
	شاخص همواری دره با وضوح مکانی بالا	Multi-Resolution Index of Valley Bottom Flatness (MrVBF)
	شدت تابش پخشیده	Diffuse insulation
	عمق دره	Valley depth
	شاخص موقعیت توپوگرافی	Topographic position index
	مساحت حوضه آبخیر	Catchment area
	طول در درجه شیب	LS-Factor
	موقعیت نسبی شیب	Relative slope position
	شاخص خیسی توپوگرافی	Topographic wetness index
	شیب	Slope
	جهت شیب	Aspect
اقلیم Climate (c)	میانگین دمای سالانه	Mean annual temperature (MAT)
	میانگین بارندگی سالانه	Mean annual precipitation
ارگانیسم Organism (o)	شاخص تفاضلی پوشش گیاهی Difference Vegetation Index	$DVI = NIR - RED$
	شاخص تفاضلی پوشش گیاهی نرمال شده Normalized Difference Vegetation Index	$NDVI = NIR - RED / NIR + RED$
	شاخص پوشش گیاهی نسبی Relative Vegetation Index	$RVI = RED - NIR$
	شاخص پوشش گیاهی تعدیل‌یافته خاک Soil adjusted Vegetation Index	$SAVI=[(NIR - R) / (NIR + R)] * (1 + S)$
	شاخص سبزینگی Greenness index	$Greenness = - 0.2848 (B2) - 0.2435(B3) - 0.5436 (B4) + 0.7243 (B5) + 0.0840(B6) - 0.1800 (B7)$
	شاخص روشنایی Brightness index	$Brightness = 0.3037 (B2) + 0.2793 (B3) + 0.4743 (B4) + 0.5585 (B4) + 0.5082 (B6) + 0.1863 (B7)$
خاک Soil (o)	احتمال حضور افق کلسیک Probability of calcic horizon	Cal.hr
	اجزا بافت خاک (رس) Soil texture component (Clay)	Clay

Bands numbers, B2: Blue, B3: Green, B4: RED, B5: NIR, B6: SWIR1, B7: SWIR2. S: عموما برابر مقدار ۰/۵ برای تعدیل اثر خاک پس‌زمینه

جنگل تصادفی

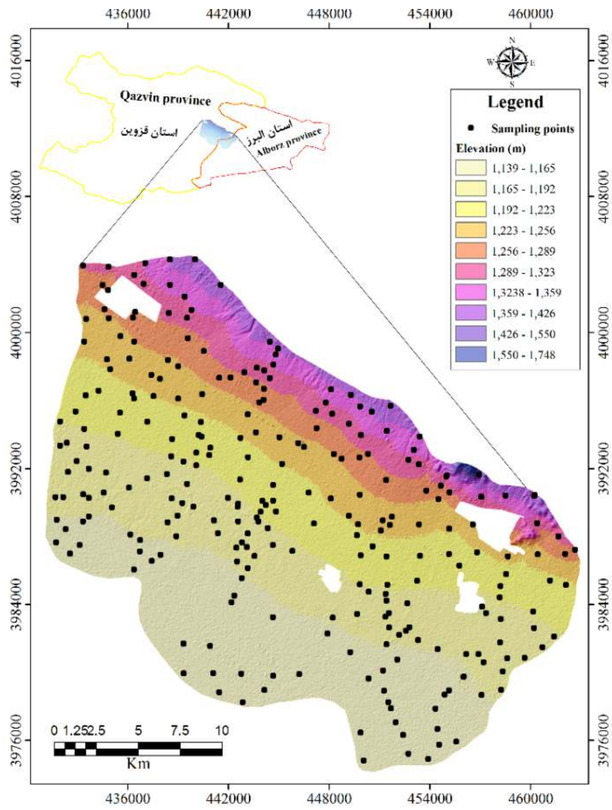
در این تحقیق از این الگوریتم جنگل تصادفی (RF) برای دست یافتن به پایدارترین نتایج در پیش‌بینی تغییرات عمودی CCE%، با تعداد درختان تصادفی ۱۰۰، ۵۰۰، ۱۰۰۰، ۱۵۰۰و ۲۵۰۰ عدد درخت با فواصل ۱۰۰ عدد و بهینه‌ترین تعداد متغیرهای کمکی در انتهای هر گره با توجه به حداقل مقدار خطا (RMSE) مورد استفاده قرار گرفت.

رگرسیون درخت تصمیم

این الگوریتم به‌عنوان یکی از الگوریتم‌های یادگیری ماشین با استفاده از مدل درختی M5 قادر است تا مقادیر متغیر هدف (CCE%) را بر اساس متغیرهای محیطی کمی در ساختار درختی ایجادشده پیش‌بینی نماید و درنهایت نتایج پیش‌بینی را به‌صورت رگرسیون خطی چند متغیره (۴۳) در برگ درختان قابل‌مشاهده است.

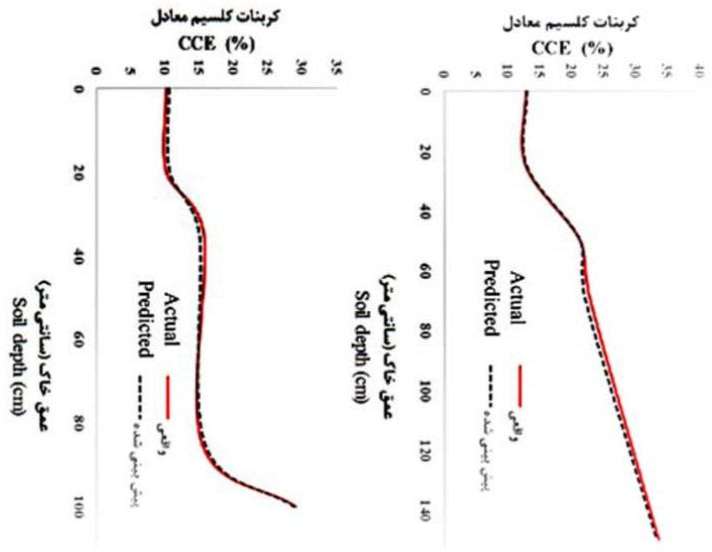
k-نزدیک‌ترین همسایگی

مدل k-نزدیک‌ترین همسایگی یکی از روش‌های غیر پارامتریک است که اساس کارکرد آن به این‌شیوه است که ابتدا فاصله اقلیدسی بین نمونه خاک موردنظر با سایر نقاط مشاهداتی محاسبه می‌گردد، سپس k تعداد از نمونه‌های مشاهداتی که در مجاورت یکدیگر قرار دارند وزن دهی می‌گردند. درنهایت بر اساس وزن هر یک از نمونه‌ها در مجموعه‌ای با k تعداد نمونه، برآوردی از داده‌های موردنظر با توجه به حداقل مقدار خطا در آن مجموعه صورت می‌پذیرد (۲۶).



شکل ۳- موقعیت منطقه مورد مطالعه و موقعیت نقاط نمونه‌برداری بر روی نقشه طبقات ارتفاعی

Figure 3- The location of study area and sampling points under classified elevation map



شکل ٤- تغییرات عمقی CCE در دو نیمرخ مشاهداتی خاک

Figure 4- Vertical variation of CCE in two example profile

اعتبارسنجی نتایج مدل‌سازی

جهت ارزیابی کارکرد الگوریتم‌های یادگیری ماشین مورداستفاده در پیش‌بینی مکانی تغییرات عمقی CCE از سه شاخص آماری ضریب تبیین (R^2)، ریشه میانگین مربعات خطا (RMSE) و میزان اریب (Bias) استفاده گردید (روابط ۱، ۲ و ۳).

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(Y_i - P_i)^2}{\sum_{i=1}^{n}(Y_i - \overline{Y})^2}$$

رابطه (۱)

$$RMSE = \sqrt{1/n\sum_{i=1}^{n}(Y_i - P_i)^2}$$

رابطه (۲)

$$Bias = \frac{\sum_{i=1}^{n} Y_i - P_i}{n}$$

رابطه (۳)

در معادلات فوق Y_i: مقادیر اندازه‌گیری شده، P_i: مقادیر پیش‌بینی‌شده CCE خاک، $\overline{Y}$ میانگین مقادیر اندازه‌گیری شده و n: تعداد مشاهدات است.

نتایج و بحث

توصیف آماری

نتایج خلاصه آماری حاکی از آن است (جدول ۲) که از سطح به عمق مقدار میانگین CCE در حال افزایش است. کمترین و بیشترین مقدار میانگین CCE به ترتیب مربوط به لایه‌های ۵-۰ و ۱۰۰-۶۰ سانتی‌متر است. حداکثر مقدار CCE نیز در عمق‌های ۳۰-۱۵ و ۱۰۰-۶۰ سانتی‌متر با مقادیر ۶۴٪ و ۶۲٪ و حداقل آن مربوط به دولایه ۵-۰ و ۶۰-۳۰ سانتی‌متر با مقادیر ۰/۲۷٪ و ۰/۳۲٪ است. با توجه به مقادیر ضریب تغییرات در لایه ۵-۰ سانتی‌متر با ۶۶/۶٪ و لایه ۶۰-۳۰ سانتی‌متر با ۵۸/۷٪ بیشترین و کمترین مقدار تغییرات کربنات کلسیم معادل در خاک‌ها مشاهده گردید. بر اساس طبقه‌بندی ارائه‌شده توسط (۴۴) CCE در تمامی عمق‌ها دارای ضریب تغییرات بالا (بیش از ۳۵٪) است. در تحقیقات سایر محققین از قبیل ژائو و همکاران (۴۶) نیز روند افزایشی در مقدار CCE از سطح تا عمق ۵۰ سانتی‌متر مشاهده گردید و در اعماق بعد تا ۲۰۰ سانتی‌متر روند نسبتاً پایداری را در مورد تغییرات عمودی CCE گزارش نمودند. به‌طور کلی خاک‌های منطقه از نظر محتوای کربنات کلسیم معادل در کلاس آهکی تا آهکی متوسط قرار دارند (۴۱). بطور کلی در بخش‌های شمالی و مرکزی منطقه مطالعاتی افزایش میزان کربنات کلسیم از سطح به عمق که عمدتاً به دلیل تأثیر فرآیند آبشویی آن‌ها از لایه‌های سطحی که منجر به انتقال و انباشت آن‌ها در لایه‌های زیرسطحی (۴۰) را می‌توان بیان نمود.

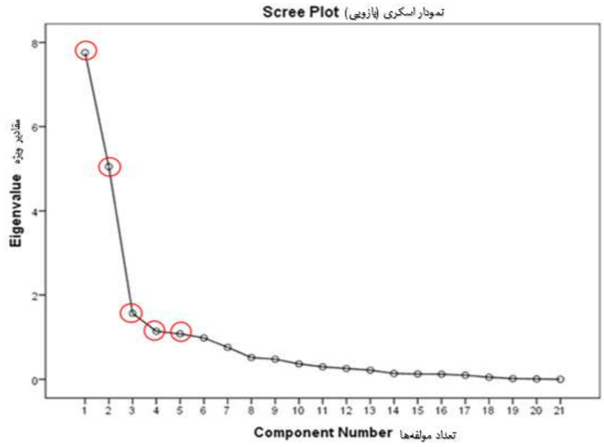
متغیرهای محیطی

نتایج روش PCA برای انتخاب مناسب‌ترین متغیرهای محیطی از میان مجموعه داده مورداستفاده نشان داد که درنهایت پنج مؤلفه اصلی از PC1 تا PC5 دارای مقادیر ارزش ویژه بالاتر از یک بودند (شکل ۵). در ادامه نتایج درصد واریانس منفرد پنج مؤلفه نخست به ترتیب نشان‌دهنده مقادیر ۲۶، ۲۰/۵، ۱۵، ۱۰ و ۹ بودند که درمجموع ۸۰/۵٪ از کل واریانس متغیرهای محیطی مورد را توجیه نمودند و درنهایت پنج متغیر کمکی Cal.hr ،Greenness ،MAT ،MrVBF و Wind Effect بالاترین ضریب ارزش ویژه را نشان دادند و درصد رس (Clay) نیز بر اساس نظر کارشناس به فرآیند مدل‌سازی اضافه گردید. متغیرهای کمکی منتخب به نحوی توسط روش PCA و نظر کارشناس انتخاب‌شده‌اند.

جدول ۲- خلاصه آماری مقادیر %CCE در اعماق استاندارد

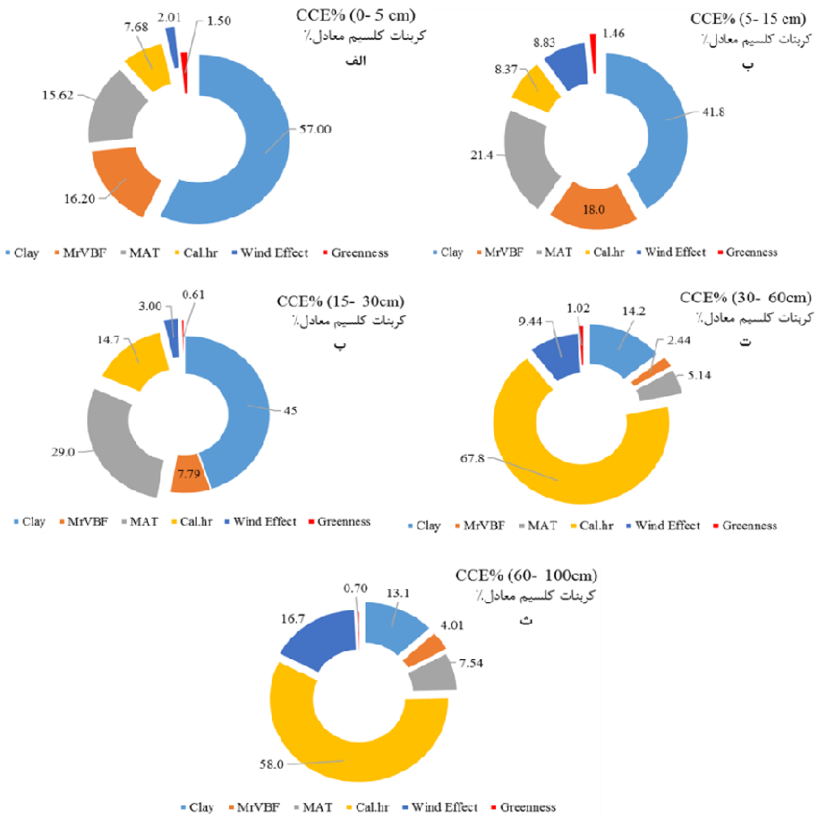
Table 2- Summary statistical of CCE% in the standardized depth

عمق خاک (سانتی‌متر) Soil depth (cm)	آموزشی Train (220)					اعتبارسنجی Test (58)				
	حداقل Min	میانگین Mean	حداکثر Max	انحراف معیار SD	ضریب تغییرات CV (%)	حداقل Min	میانگین Mean	حداکثر Max	انحراف معیار SD	ضریب تغییرات CV (%)
0-5	0.27	12.0	32	8.0	66.6	0.30	11.8	52	7.50	63.5
5-15	0.50	12.2	56	8.02	65.4	0.55	12.0	50	7.55	62
15-30	0.45	13.0	64	8.1	62.3	0.40	12.2	58	7.8	63
30-60	0.32	15.5	60	9.1	58.7	0.35	15.1	62	9.2	60
60-100	0.41	16.2	62	10.3	63.5	0.43	15.9	64	9.81	61



شکل ۵- نمودار اسکری (بازوئی) برای تعیین تعداد مولفه‌های مناسب

Figure 5- The scree plot for determining the principal component number



شکل ٦- اهمیت نسبی متغیرهای محیطی در پیش‌بینی تغییرات عمقی CCE%

Figure 6- The relative importance of environmental covariates in prediction of CCE % depth variation

رحمانی و همکـاران (۳۱) نیـز از روش PCA اسـتفاده نمودنـد و مشابه با نتایج این تحقیق پنج مؤلفه نخست در این روش حدود ۸۰/۲ درصد از واریانس مجموع را به خود اختصـاص داد، در همـین ارتبـاط امیریان چکـان و همکـاران (۵) روش PCA را بـرای انتخـاب بهینـه متغیرهای محیطی در پیش‌بینی ویژگی‌های فیزیکـی خـاک پیشـنهاد نمودند. نتایج اهمیت نسبی[1] متغیرهای محیطی پـیش‌بینـی کننـده در

1- Relative importance

اعماق استاندارد مورد بررسی در شکل ۶ نشان داده شده است.

بر اساس شکل‌های (۶ الف، ب و پ) متغیر خاک، Clay به ترتیب با توجیه ۵۷، ۴۱/۸ و ۴۵ درصد از کل واریانس تغییرات مکانی CCE را در سه عمق استاندارد ۵-۰ ۱۵-۵ و ۳۰-۱۵ سانتی‌متر بالاترین اهمیت نسبی را نسبت به سایر متغیرهای کمکی ارائه نمود. که مشاهده این نتایج بر اساس (شکل ۷ الف، ۷ب و ۷پ) که پراکنش مکانی آهک را به‌ترتیب در سه عمق استاندارد۳۰-۱۵، ۱۵-۵ و ۵-۰ سانتی‌متر را نشان می‌دهند با جدول همبستگی (جدول ۳) که نشان دهنده همبستگی مثبت و معنی‌دار درصد کربنات کلسیم معادل در این اعماق با درصد رس می‌باشد، که این نتایج با یافته‌های (۱۱) نیز همخوانی دارد. بعد از Clay در دو عمق اول (۵-۰ و ۱۵-۵) متغیر توپوگرافی MrVBF با توجیه به ترتیب ۱۸ و ۱۶ درصد از تغییرات CCE نسبت به سایر متغیرهای محیطی در درجه دوم اهمیت مشاهده گردید. در همین راستا موسوی و همکاران (۲۲) در مدل‌سازی مکانی کربنات کلسیم معادل و تهیه نقشه پیش‌بینی آن در خاک سطحی اراضی قروه-دهگلان مشاهده نمودند که متغیرهای محیطی توپوگرافی از قبیل DEM، شاخص موقعیت توپوگرافی و فاصله اقلیدسی از شبکه آبراهه بالاترین میزان اهمیت را دارا بودند. در اعماق استاندارد بعدی ۶۰-۳۰ و ۱۰۰-۶۰ سانتی‌متر نتایج بیانگر بارزتر شدن نقش متغیر محیطی Cal.hr نسبت به سایر متغیرهای محیطی با ارائه ۶۷/۸ و ۵۸ درصد از کل واریانس تغییرات عمقی CCE بود؛ بنابراین به نظر می‌رسد برخلاف نتایج مولدر و همکاران (۲۴) که معتقدند از سطح به عمق به دلیل کاهش همبستگی فاکتورهای محیطی با ویژگی‌های خاک، دقت نقشه‌های پیش‌بینی کاهش می‌یابد، اما در این تحقیق به دلیل وجود متغیر کمکی Cal.hr که برگرفته از خود خاک است، موجب افزایش دقت مدل‌سازی در اعماق زیرسطحی گردید و به نحوی کاهش تأثیر متغیرهای محیطی در توجیه تغییرات عمقی ویژگی‌های خاک پوشش داده است. در پژوهش دیگری نیز نتایج محققین بیانگر اهمیت و تأثیر بیشتر متغیرهای کمکی (رس، کربن آلی) برگرفته از خاک، در نقشه‌برداری رقومی ویژگی‌های خاک نسبت به شاخص‌های سنجش دور بود (۴). همان‌گونه که مشاهده می‌گردد (شکل ۶) شاخص سنجش‌ازدوری سبزینگی (Greenness) به‌طور کلی قادر به توجیه واریانس قابل‌توجهی از تغییرات CCE% نبود.

برخلاف نتایج تحقیق حاضر در مطالعه‌ای زراعت پیشه و همکاران (۴۰) همبستگی بالایی را بین متغیرهای سنجش‌ازدور پوشش گیاهی و CCE گزارش کردند. روند اهمیت نسبی پارامتر اقلیمی MAT از لایه اول تا لایه استاندارد سوم افزایش یافت، به نحوی‌که بیشترین تأثیر آن در عمق ۳۰-۱۵ سانتی‌متر بود. متغیر کمکی Wind Effect در لایه‌های زیرین ۶۰-۳۰ و ۱۰۰-۶۰ سانتی‌متر نسبت به سه لایه‌ی استاندارد فوقانی اهمیت بیشتری را نشان داد. چنین به نظر می‌رسد، شاخص تاثیر باد در دو عمق آخر به دلیل تاثر باد بر افزایش و یا کاهش میزان تبخیر و تعرق و تاثیرگذاری غیر مستقیم بر میزان انحلال و یا صعود موئینگی کربنات کلسیم معادل از پایین به بالا سبب تجمع کربنات کلسیم در اعماق مشاهده شده برای افق کلسیک گشته است هر چند که در این اعماق پارامتر دارای اهمیت بالا با بیش از ۶۰ درصد هنوز پارامتر Cal.hr می‌باشد و در درجه بعدی پارامتر اثر باد دارای درجه اهمیت نسبی بیش از ۱۵ درصد است. پارامترهای اقلیمی، توپوگرافی، مواد مادری و ارگانیسم به‌عنوان فاکتورهای محیطی کنترل‌کننده تغییرات ویژگی‌های خاک می‌باشند (۱۶). به‌طور کلی نتایج اهمیت نسبی متغیرهای پیش‌بینی کننده CCE در لایه‌های سطحی و زیرسطحی نشان داد که تغییرات مکانی CCE به مقدار قابل‌توجهی در منطقه مورد مطالعه توسط رس و در اعماق زیرسطحی به‌وسیله نقشه احتمال افق مشخصه کلسیک قابل توجیه است که این نتیجه نشان از تأثیر بالای فاکتور "s" بر نتایج نقشه‌برداری رقومی ویژگی‌های خاک است.

ارزیابی کارایی الگوریتم‌های یادگیری ماشین

نتایج کارایی الگوریتم‌های یادگیری ماشین RF، DTr و k-NN در پیش‌بینی تغییرات مکانی CCE از سطح به عمق در جدول ۴ ارائه‌شده است. الگوریتم RF بر اساس آماره‌های R^2 و RMSE دارای بالاترین میزان دقت و حداقل خطا بود. به‌نحوی‌که مقدار آماره R^2 در الگوریتم RF از ۰/۷۶ تا ۰/۸۳ و مقادیر RMSE از ۳/۲۱٪ در لایه ۵-۰ سانتی‌متر تا ۱۴٪ در لایه ۱۰۰-۶۰ سانتی‌متر متغیر است. به‌طور کلی بر مبنای RF از سطح به عمق با افزایش R^2، مقدار RMSE یک‌روند کاهشی را نشان می‌دهد. در حالی‌که مقدار Bias در امتداد پروفیل خاک یک‌روند نامنظم را نشان داد.

جدول ۳- ضریب همبستگی پیرسون بین درصد رس و کربنات کلسیم معادل در پنج عمق استاندارد (عمق بر حسی سانتی‌متر)

Table 3- Pearson correlation coefficient between Clay and CCE in five standard depths (cm)

Clay (60-100)	Clay (30-60)	Clay (15-30)	Clay (5-15)	Clay (0-5)	متغیرهای خاک
				0.22**	CCE (0-5)
			0.23**		CCE (5-15)
		0.25**			CCE (15-30)
	0.23**				CCE (30-60)
0.23**					CCE (60-100)

جدول ٤- نتایج اعتبارسنجی الگوریتم‌های پیش‌بینی کننده %CCE در اعماق استاندارد
Table 4- Validation results of CCE% prediction algorithms in standard depths

عمق خاک (سانتی‌متر) Soil depth (cm)	الگوریتم یادگیری ماشین (MLA)	شاخص‌های اعتبارسنجی: ضریب تبیین R^2	جذر میانگین مربعات خطا RMSE (%)	اریب Bias
0-5	جنگل تصادفی (RF)	0.82	3.58	0.06
5-15		0.85	3.46	0.16
15-30		0.87	2.69	0.14
30-60		0.89	2.34	0.29
60-100		0.90	2.14	0.33
0-5	رگرسیون درخت تصمیم (DTr)	0.39	5.14	0.35
5-15		0.41	5.05	0.32
15-30		0.48	4.23	0.28
30-60		0.51	4.01	0.25
60-100		0.52	3.91	-0.21
0-5	k- نزدیک‌ترین همسایه (k-NN)	0.37	5.18	0.29
5-15		0.29	5.96	0.53
15-30		0.39	5.10	-0.46
30-60		0.21	7.98	0.62
60-100		0.15	9.22	1.62

در مدل DTr باوجود اینکه مقادیر دقت کمتر و خطای پیش‌بینی بیشتری برای CCE حاصل گردید اما روند مشابهی با RF نشان داد. نتایج کارایی این الگوریتم با مقادیر R^2 ۰/۳۹ تا ۰/۵۲ و RMSE ۳/۹۱٪ تا ۵/۱۴٪ متغیر بود (جدول ۲). برخلاف دو الگوریتم (RF و DTr) در مورد k-NN نتایج اعتبارسنجی آن به‌جز در لایه ۳۰-۱۵ سانتی‌متر، در کلیه لایه‌های استاندارد روند کاملاً متفاوتی را نشان داد، بطوری‌که از سطح به عمق با کاهش R^2 مقادیر RMSE افزایش یافت؛ بنابراین از نظر میزان دقت سه الگوریتم پیش‌بینی کننده مکانی به‌ترتیب RF دارای دقت پیش‌بینی خوب[1]، DTr متوسط[2] و k-NN ضعیف[3] می‌باشند (۴۲). در مطالعه مقایسه‌ای که در مرکز ایران انجام گردید دو مدل کوبیست (CB) و سپس RF را در پیش‌بینی مکانی CCE در خاک سطحی به عنوان مدل برتر معرفی نمودند (۴۵). رحمانی و همکاران (۳۱) توانائی بالاتر مدل RF در مقایسه با روش SoLIM با مقادیر R^2 برابر ۰/۶۳ و ۰/۴۲ در پیش بینی CCE سطحی گزارش نمودند. همچنین در پژوهشی در غرب ایران، مدل RF با مقادیر (R^2=۰/۴ و RMSE=۱۱/۶٪) نسبت به دو روش کریجینگ معمولی و وزن‌دهی معکوس فاصله در پیش‌بینی CCE (عمق ۳۰-۰ سانتی‌متر) به عنوان مدل برتر گزارش گردید (۲۱).

تغییرپذیری مکانی کربنات کلسیم معادل

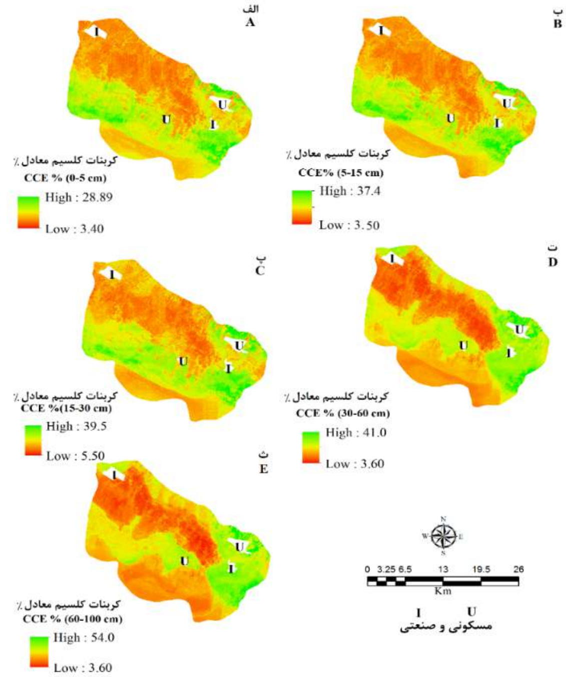
با توجه به اینکه الگوریتم RF نسبت به دو روش دیگر بالاترین میزان صحت را ارائه نمود نقشه‌های نهایی پیش‌بینی مکانی CCE در پنج عمق استاندارد بر اساس آن ارائه‌شده است (شکل ۷). روند کلی تغییرات CCE در سه لایه فوقانی خاک ۳۰-۱۵ و ۱۵-۵، ۵-۰ سانتی‌متر به یکدیگر مشابه است در حالی‌که با افزایش میزان عمق تغییرات توزیع مکانی آهک به‌ویژه در بخش‌های شمالی نشان‌دهنده افزایش میزان CCE است. بر اساس شکل‌های (۷ الف و ۷ ب و ۷ پ) که روندهای مشابه‌تری با یکدیگر دارند، بیشترین مقدار CCE در یک امتداد از شرق منطقه به غرب منطقه است. همان‌طور که در نتایج قبل توضیح داده شد، پیش‌ران‌های اصلی تغییرات CCE در لایه‌های سطحی Clay و دو پارامتر توپوگرافی MrVBF و Wind Effect بوده است. در همین راستا کشاورزی و همکاران (۱۱) همبستگی مثبت و معنی‌داری (در سطح ۱ درصد) بین درصد رس و کربنات کلسیم معادل در منطقه زیاران دشت قزوین گزارش نمودند. با توجه به اینکه شکل غالب کربنات کلسیم مشاهده شده در خاک‌های منطقه از نوع ثانویه می‌بودند، این فرم از کربنات عمدتا در خاکدانه‌های با اندازه رس و سیلت تشکیل می‌شوند (۱۲). در مطالعه‌ی دیگری محققین در مورد ارتباط رس و کربنات کلسیم معتقدند که در خاک‌هایی که مقدار فراوانی کربنات وجود دارد آنها با هم‌آوری رس در افق‌های فوقانی منجر به افزایش اندازه آن شده و در عمل از جابه‌جایی آن جلوگیری می‌کنند (۳). در منطقه پیرانشهر آذربایجان غربی نیز رضاپور (۳۲) مقادیر بالاتر کربنات کلسیم معادل را در افق سطحی خاک‌های ورتی‌سول نسبت به کمبی‌سول‌های تحت کاربری‌های زراعی و مرتعی در طی مطالعه خود گزارش نمود. بخش‌های شمالی منطقه دارای بیشترین میزان ارتفاع می‌باشند (شکل ۴) که به دلیل وجود ارتباط بین میزان ارتفاع، بارش و رژیم رطوبتی خاک (زریک) و از طرفی بر اساس اطلاعات خاکشناسی این بخش‌ها دارای بافت سبک سطحی (لوم شنی)، زهکشی مناسب و درصد رس

1- Well
2- Moderate
3- Weak

کمتر (شکل ۳ ب) منجر به شستشو و انتقال کربنات کلسیم از لایه‌های سطحی (۰–۳۰ سانتی‌متر) به اعماق زیرین شده است که این نتایج با مشاهدات (۳۰ و ۱) مطابقت دارد. در بخش‌های میانی منطقه به سمت شمال غرب به دلیل وجود کاربری زراعت آبی متمرکز، آبیاری طولانی‌مدت توسط کشاورزان و بافت سطحی متوسط (لوم) تا سبک (لوم شنی) کربنات کلسیم از سطح شسته و در اعماق پایین‌تر تجمع یافته است که شرایط این بخش از منطقه بر اساس کلیه نقشه‌های پیش‌بینی تا عمق ۱۰۰ سانتی‌متری CCE در حداقل میزان خود است (۸). دلیل پایین بودن مقدار CCE در بخش‌های جنوبی منطقه عمدتاً ناشی از حضور گچ و شوری بالا و نوع مواد مادری از نوع (سازند کفه‌های گلی نمکی) است که ازنظر روند تکامل به دلیل میزان حلالیت بالاتر املاح و گچ نسبت به کربنات کلسیم موجب شده که عمق تجمع CCE خارج از حداکثر میزان عمق مورد مطالعه باشد (۸). در مورد اعماق زیرین ۶۰–۳۰ و ۱۰۰–۶۰ سانتی‌متر بر اساس نتایج اهمیت نسبی متغیر Cal.hr بیشترین میزان اهمیت را نسبت به سایر متغیرها داشت که موجب شده تا روند تغییرات زیرسطحی آهک به شکل قابل‌توجهی متأثر از تغییرات این ویژگی که از جنس خود خاک نیز است، مشاهده گردد. بر اساس شکل‌های (۷ ت و ۷ ث) در بخش‌های شرقی و شمال شرقی منطقه به دلیل وجود بافت خاک زیرسطحی سنگین‌تر (لوم رسی) نسبت به لایه‌های فوقانی به همراه آبشویی کمتر شرایط مناسبی برای تجمع کربنات کلسیم فراهم‌شده به‌ویژه مقادیر CCE در این مناطق به‌طور متوسط بیش از ۴۰٪ است. به‌طورکلی وضعیت خاک‌های منطقه در اعماق مورد پژوهش نشان می‌دهد که در لایه‌های سطحی ۳۰–۱۵ و ۱۵–۵، ۵–۰ سانتی‌متر حدود ۶۰٪ از خاک‌ها در کلاس آهکی ضعیف تا آهکی و مابقی آهکی متوسط تا آهکی شدید می‌باشند، درحالی‌که در اعماق زیرین ۶۰–۳۰ و ۱۰۰–۶۰ سانتی‌متر بیش از ۵۰ درصد منطقه در کلاس آهکی شدید، ۳۰٪ درصد در کلاس آهکی متوسط و حدود ۲۰٪ در کلاس آهکی و آهکی ضعیف قرار دارند (۱۹).

نتیجه‌گیری

پژوهش حاضر باهدف بررسی کارایی سه الگوریتم یادگیری ماشین RF، DTr و k-NN در مدل‌سازی و پیش‌بینی تغییرات مکانی CCE در پنج عمق استاندارد ۱۰۰–۶۰، ۶۰–۳۰، ۳۰–۱۵، ۱۵–۵ و ۵–۰ سانتی‌متر، متناظر با پروژه جهانی تهیه نقشه خاک اجرا گردید. به‌طور کلی نتایج استفاده از تابع اسپیلاین ارائه یک تخمین قابل‌قبول از مقادیر CCE در هر یک از اعماق مورد بررسی با حداقل خطا و حداکثر انطباق میان داده‌های واقعی و پیش‌بینی‌شده که می‌توان نتایج آن را برای همسازی ویژگی‌های خاک به‌ویژه CCE در سایر مناطق خشک و نیمه‌خشک، مشابه با شرایط منطقه مورد مطالعه استفاده نمود.



شکل ۷- نقشه‌های رقومی تغییرات عمقی CCE با استفاده از مدل RF

Figure 7- Digital maps of CCE vertical variation by RF model

روش انتخاب متغیر PCA به‌خوبی توانست با تعیین پنج مؤلفه اصلی و انتخاب یک ویژگی با بالاترین ضریب ارزش ویژه در هر مؤلفه بیش از ۸۰٪ واریانس تجمعی متغیرهای محیطی موجود را تأمین نماید و همچنین پوشش کاملی از فاکتورهای مدل "scorpan" شامل توپوگرافی (Wind Effect ،MrVBF)، ارگانیسم (Greenness)، اقلیم (MAT) و خاک (Cal.hr) را ارائه نماید. الگوریتم یادگیری ماشین RF نسبت به دو الگوریتم دیگر (DTr و k-NN) بیشترین میزان دقت و حداقل خطا را در پیش‌بینی مکانی CCE در همه اعماق نشان داد؛ بنابراین استفاده از الگوریتم RF با ارائه نتایج قابل‌اعتماد برای تهیه نقشه تغییرات سطحی و عمقی ویژگی‌های خاک توصیه می‌گردد. پراکنش مکانی CCE در سه عمق استاندارد ۳۰-۱۵ و ۱۵-۵، ۵-۰ سانتی‌متر تا حد بسیار زیادی متأثر از تغییرات درصد رس در منطقه بود که همبستگی بالایی را با CCE نشان داد، درحالی‌که مقادیر CCE در اعماق ۱۰۰-۶۰، ۶۰-۳۰ سانتی‌متر متأثر از نقشه احتمال افق کلسیک (Cal.hr) به‌عنوان مهم‌ترین پیش‌ران تغییرات این ویژگی بود که نشان از غالبیت فرآیندهای ژنتیکی خاک و تأثیر بیشتر آن‌ها در توجیه تغییرات مکانی CCE نسبت به سایر فاکتورهای خاکسازی مورد استفاده در این اعماق بود. با توجه به اینکه به‌طور عمده بخش‌های شمال شرق و میانی منطقه در کلاس آهکی شدید (۵۰-۲۵ درصد) قرار می‌گیرند، بنابراین به کارشناسان، مدیران و بهره‌برداران اراضی توصیه می‌گردد تا اقدامات اصلاحی از قبیل استفاده از کودهای آلی (حیوانی، کمپوست و ورمی‌کمپوست) و کودهای شیمیای با پایه ترکیبات سولفاته (سولفات آمونیوم) جهت تعدیل اثرات ناشی از حضور آهک در خاک مدنظر قرار دهند، و همچنین بهبود شرایط شیمیایی خاک را فراهم می‌نمایند و منجر به جذب بهتر عناصر غذایی ضروری پرمصرف مثل فسفر و کم مصرف (از قبیل روی، مس، آهن و منگنز) می‌انجامد، که در ادامه موجبات ارتقا حاصلخیزی خاک‌ها و افزایش عملکرد محصولات واقع در الگوی کشت منطقه گردد.

سپاسگزاری

از کلیه مدیران شهرستان و کارشناسان بخش کشاورزی نظرآباد جهت همکاری در انجام مراحل مطالعات میدانی و نمونه‌برداری خاک‌ها در راستای تکمیل هر چه بهتر این پژوهش همکاری نمودند، کمال تشکر و قدردانی را داریم.

منابع

1- Amirian C.A., Taghizadeh Mehrjardi R., Sarmadian F., and Mohammadi J. 2018. Study of lateral and vertical distribution of soil calcium carbonate using geostatistics and spline functions. (In Persian with English abstract)

2- Arrouays D., Grundy M.G., Hartemink A.E., Hempel J.W., Heuvelink G.B., Hong S.Y., Lagacherie P., Lelyk G., McBratney A.B., McKenzie N.J., and dL Mendonca-Santos M. 2014. GlobalSoilMap: Toward a fine-resolution global grid of soil properties. Advances in Agronomy 125: 93-134.

3- Asgari Hafshejani N., and Jafari S. 2017. The study of particle size distribution of calcium carbonate and its effects on some soil properties in Khuzestan province. Iran Agricultural Research 36(2): 71-80.

4- Bouslihim Y., Rochdi A., and Paaza N.E.A. 2021. Machine learning approaches for the prediction of soil aggregate stability. Heliyon 7(3): e06480.

5- Chakan A.A., Taghizadeh-Mehrjardi R., Kerry R., Kumar S., Khordehbin S., and Khanghah S.Y. 2017. Spatial 3D distribution of soil organic carbon under different land use types. Environmental Monitoring and Assessment 189(3): 131.

6- Esfandiarpour Boroujeni I., ShahiniShamsabadi M., Shirani H., Mosleh Z., BagheriBodaghabadi M., and Salehi M.H. 2020. Assessment of different digital soil mapping methods for prediction of soil classes in the Shahrekord plain, Central Iran. Catena 193: 104648.

7- Esmaeili E., Shahbazi F., Sarmadian F., Jafarzadeh A.A., and Hayati B. 2021. Land capability evaluation using NRCS agricultural land evaluation and site assessment (LESA) system in a semi-arid region of Iran. Environmental Earth Sciences 80(4): 1-14.

8- FAO. 1973. Irrigation, Drainage and salinity. FAO/UNESCO.

9- Hengl T., Mendes de Jesus J., Heuvelink G.B., Ruiperez Gonzalez M., Kilibarda M., Blagotić A., Shangguan W., Wright M.N., Geng X., Bauer-Marschallinger B., and Guevara M.A. 2017. SoilGrids250m: Global gridded soil information based on machine learning. PLoS one 12(2): e0169748.

10- Hengl T., Miller M.A., Krizan J., Shepherd K.D., Sila A., Kilibarda M., Antonijevic O., Glušica L., Dobermann A., Haefele S.M., and McGrath S.P. 2021. African soil properties and nutrients mapped at 30 m spatial resolution using two-scale ensemble machine learning. Scientific Reports 11(1): 1-18.

11- Keshavarzi A., Sarmadian F., Labbafi R., and Ahmadi A. 2011. Developing pedotransfer functions for estimating field capacity and permanent wilting point using fuzzy table look-up scheme. Computer and Information Science 4(1): 130.

12- Khodaverdiloo H., Homaee M., van Genuchten M.T., and Dashtaki S.G. 2011. Deriving and validating pedotransfer functions for some calcareous soils. Journal of Hydrology 399(1-2): 93-99.
13- Khaledian Y., and Miller B.A. 2020. Selecting appropriate machine learning methods for digital soil mapping. Applied Mathematical Modelling 81: 401-418.
14- Kuhn M., and Johnson K. 2013. Applied predictive modeling (Vol. 26, p. 13). New York: Springer.
15- Lacoste M., Minasny B., McBratney A., Michot D., Viaud V., and Walter C. 2014. High resolution 3D mapping of soil organic carbon in a heterogeneous agricultural landscape. Geoderma 213: 296-311.
16- McBratney A.B., Santos M.M., and Minasny B. 2003. On digital soil mapping. Geoderma 117(1-2): 3-52.
17- Mahmoudabadi E., Karimi A., Haghnia G.H., and Sepehr A. 2017. Digital soil mapping using remote sensing indices, terrain attributes, and vegetation features in the rangelands of northeastern Iran. Environmental Monitoring and Assessment 189(10): 1-20.
18- Malone B.P., McBratney A.B., Minasny B., and Laslett G.M. 2009. Mapping continuous depth functions of soil carbon storage and available water capacity. Geoderma 154(1-2): 138-152.
19- McDonald R.C., Isbell R.F., Speight J.G., Walker J., and Hopkins M.S. 1998. Australian soil and land survey: field handbook (No. Ed. 2). CSIRO publishing.
20- Mosleh Z., Salehi M.H., Jafari A., Borujeni I.E., and Mehnatkesh A. 2016. The effectiveness of digital soil mapping to predict soil properties over low-relief areas. Environmental Monitoring and Assessment 188(3): 195.
21- Mousavi S.R., Parsayi F., Rahmani A., Sedri, M.H., and Kohsar Bostani M. 2020. Spatial Prediction Some of the Surface Soil Properties Using Interpolation and Machine Learning Models. Journal of Soil Management and Sustainable Production 10(3): 27-49. (In Persian with English abstract).
22- Mousavi S.R., Sarmadian F., Dehghani S., Sadikhani M.R., and Taati A. 2017. Evaluating inverse distance weighting and kriging methods in estimation of some physical and chemical properties of soil in Qazvin Plain. Eurasian Journal of Soil Science 6(4): 327-336.
23- Mousavi S.R, Sarmadian F., Omid M., and Bogaert P. 2021. Digital modeling of three-dimensional soil salinity variation using machine learning algorithms in arid and semi-arid land of Qazvin plain. Iranian Journal of Soil and Water Research, doi: 10.22059/ijswr.2021.323030.668957. (In Persian with English abstract)
24- Mulder V.L., Lacoste M., Richer-de-Forges A.C., Martin M.P., and Arrouays D. 2016. National versus global modelling the 3D distribution of soil organic carbon in mainland France. Geoderma 263: 16-34.
25- Nelson R.E. 1982 Carbonate and gypsum. In: Page AL (ed) Methods of soil analysis. American Society of Agronomy, Madison, pp 181–197.
26- Nemes A., Rawls W.J., and Pachepsky Y.A. 2006. Use of the nonparametric nearest neighbor approach to estimate soil hydraulic properties. Soil Science Society of America Journal 70(2): 327-336.
27- Pahlavan-Rad M.R., and Akbarimoghaddam A. 2018. Spatial variability of soil texture fractions and pH in a flood plain (case study from eastern Iran). Catena 160: 275-281.
28- Padarian J., Minasny B., and McBratney A.B. 2019. Using deep learning for digital soil mapping. Soil 5: 79–89.
29- Parsaie F., Firouzi A.F., Mousavi S.R., Rahmani A., Sedri M.H., and Homaee M. 2021. Large-scale digital mapping of topsoil total nitrogen using machine learning models and associated uncertainty map. Environmental Monitoring and Assessment 193(4): 1-15.
30- Presley D.R., Ransom M.D., Kluitenberg G.J., and Finnell P.R. 2004. Effects of thirty years of irrigation on the genesis and morphology of two semiarid soils in Kansas.
31- Rahmani A., Sarmadian F., Mousavi S.R., and Khamoshi S.E. 2020. Application of Geomorphometric attributes in digital soil mapping by using of machine learning and fuzzy logic approaches. Journal of Range and Watershed Managment 73(1): 105-124. (In Persian)
32- Rezapour S. 2014. Response of some soil attributes to different land use types in calcareous soils with Mediterranean type climate in north-west of Iran. Environmental Earth Sciences 71(5): 2199-2210.
33- Rossel R.V., Chen C., Grundy M.J., Searle R., Clifford D., and Campbell P.H. 2015. The Australian three-dimensional soil grid: Australia's contribution to the GlobalSoilMap project. Soil Research 53(8): 845-864.
34- Rostaminia M., Nouri N., Keshavarzi A., and Rahmani A. 2019. Quantitative Evaluation and Zoning of Spatial Distribution of Soil Quality Index in Some Parts of Arid and Semi-Arid Lands of Western Iran (Case Study: Kane Sorkh Region, Ilam Province). Iranian Journal of Soil and Water Research 50(7): 1701-1719. (In Persian with English abstract)
35- Sreenivas K., Dadhwal V.K., Kumar S., Harsha G.S., Mitran T., Sujatha G., Suresh G.J.R., Fyzee M.A., and Ravisankar T. 2016. Digital mapping of soil organic and inorganic carbon status in India. Geoderma 269: 160-173.
36- Staff S.S. 2014. Keys to Soil Taxonomy, 12th Edn Washington. DC: Natural Resources Conservation Service, United States Department of Agriculture.
37- Taghizadeh Mehrjardi R., Nabiollahi K., and Kerry R. 2016. Digital mapping of soil organic carbon at multiple depths using different data mining techniques in Baneh region, Iran. Geoderma 266: 98-110.
38- Taghizadeh Mehrjardi R., Minasny B., Sarmadian F., and Malone P.B. 2014a. Digital mapping of soil salinity in Ardakan region, central Iran. Geoderma 213: 15-28.

39- Taghizadeh-Mehrjerdi R., Amirin Chakan A., and Sarmadian F. 2014b. 3D digital mapping of soil cation exchange capacity in Dorud, Lorestan province. Journal of Water and Soil 28: 998-1010. (In Persian with English abstract)
40- Tan W.F., Zhang R., Cao H., Huang C.Q., Yang Q.K., Wang M.K., and Koopal L.K. 2014. Soil inorganic carbon stock under different soil types and land uses on the Loess Plateau region of China. Catena 121: 22-30.
41- Vargas R., Pankova E.I., Balyuk S.A., Krasilnikov P.V., and Khasankhanova G.M. 2018. Handbook for saline soil management. FAO/LMSU.
42- Viscarra Rossel R.A., and McBratney A.B. 2008. Diffuse reflectance spectroscopy as a tool for digital soil mapping. In 'Digital soil mapping with limited data'. Developments in Soil Science series. (Eds AE Hartemink, AB McBratney, L Mendonça-Santos) (Elsevier Science: Amsterdam).
43- Wang Y., and Witten I.H. 1997. Inducing model trees for continuous classes. In Proceedings of the Ninth European Conference on Machine Learning. pp. 128–137.
44- Wilding L.P. 1985. Spatial variability: its documentation, accomodation and implication to soil surveys. In Soil spatial variability, Las Vegas NV, 30 November-1 December 1984 (pp. 166-194).
45- Zeraatpisheh M., Ayoubi S., Jafari A., Tajik S., and Finke P. 2019. Digital mapping of soil properties using multiple machine learning in a semi-arid region, central Iran. Geoderma 338: 445-452.
46- Zhao W., Zhang R., Huang C., Wang B., Cao H., Koopal L.K., and Tan W. 2016. Effect of different vegetation cover on the vertical distribution of soil organic and inorganic carbon in the Zhifanggou Watershed on the loess plateau. Catena 139: 191-198.

Journal of Water and Soil
Vol. 35, No. 5, Dec.-Jan. 2021, p. 719-734

نشریه آب و خاک (علوم و صنایع کشاورزی)
جلد ۳۵، شماره ۵، آذر – دی ۱۴۰۰، ص. ۷۳۴-۷۱۹

Modeling the Vertical Soil Calcium Carbonate Equivalent Variation by Machine Learning Algorithms in Qazvin Plain

S.R. Mousavi[1]- F. Sarmadian[2*]- M. Omid[3]- P. Bogaert[4]
Received: 16-08-2021
Accepted: 25-09-2021

Introduction: Calcium Carbonate Equivalent (CCE) is one of the key soils properties in arid and semi-arid regions. The study of spatial variability of surface and subsurface layers is important in the sustainable land management of arable soils. This study aimed to model the spatial distribution of CCE percentage by using three machine learning algorithms including Random Forest (RF), Decision Tree regression (DTr) and k-Nearest Neighbor (k-NN) at five standard depths of 0-5, 5-15, 15-30, 30-60, and 60-100 cm.

Material and Methods: The study area with 60,000 ha includes the major part of the lands of Qazvin plain located on the border of Qazvin and Alborz provinces. Field and laboratory surveys included 278 representative profiles were excavated, described by the horizon, and determined physicochemical properties. The studied soils have a very high diversity in soil moisture (Aridic, Xeric, and Aquic) and temperature regimes (Thermic). These variations have led to the formation of eight great groups of soils in the region based in the *USDA* soil classification system with the three classes of *Haploxerepts*, *Calcixerepts*, and *Haplocalcids* were the dominant soil classes in the study area. A total of 22 environmental covariates, including 12 variables extracted from the primary and secondary derivation of digital elevation model (DEM), six remote sensing (RS) indicators, two climatic parameters, and two soil covariates were prepared, and then the most appropriate environmental covariates were selected using principal component analysis (PCA) and expert knowledge. The CCE percentage data were randomly divided into two parts, 80% for training and 20% for testing, which was then modeled by three machine learning algorithms RF, DTr, and k-NN, and were evaluated by some statistical indices as coefficient determination (R^2), root mean square error (RMSE) and Bias.

Results and Discussion: The results of harmonizing the CCE values at the genetic horizons with the standard depths showed the high efficiency of the spline depth function in providing an acceptable estimate with minimum error and maximum agreement between observed and predicted values. The PCA method showed that the first to fifth components with the explanation of more than 80% of cumulative variance were Multi-Resolution Index of Valley Bottom Flatness (MrVBF), Mean Annual Temperature (MAT), Greenness index (Greenness), Probability of Calcic horizon (Cal.hr), and Wind Effect environmental covariates which had the highest eigenvalues. Besides, Clay was selected on expert knowledge-based. The relative importance (RI) of the environmental covariates showed the spatial distribution of CCE were affected by Clay with an explanation of more than 57%, 41.8% and 45% of its variance at three surface depths of 0-5, 5-15, and 15-30 cm, while the Cal.hr covariate had the highest impact in the spatial prediction of CCE compared to other predictors as auxiliary variables with 67.8% and 52.8% justification, respectively, at two depths of 30-60 and 60-100 cm. Hence, using the calcic horizon probability Map (Cal.hr) as a derivative soil factor made it possible to produce more appropriate final maps, while preventing the reduction of the accuracy of the modeling results in the subsoils. The auxiliary variable of remote sensing, i.e., Greenness, could not show a significant impact on the expression of the variation of CCE percentage at all studied depths. Unlike remote sensing indices, the topographic attribute of the MrVBF, at two standard depths of 0-5 and 5-15 cm, the MAT at a depth of 15-30 cm, and the Wind Effect at the standard depths 30-60 and 60-100 cm, after the soil covariates, were the most effective in justifying the spatial variations of CCE%. RF algorithm with a range of R^2 values of 0.83 - 0.76 and RMSE of 2.14% - 2.21%

1 and 2- Ph.D. Studunet of Soil Resource Management and Professor, Department of Soil Science and Engineering, College of Agriculture and Natural Resources, University of Tehran, Karaj, Iran, respectively.
(*- Corresponding Author Email: fsarmad@ut.ac.ir)
3- Professor of Agricultural Machinery Engineering Department, College of Agriculture and Natural Resources, University of Tehran, Karaj, Iran
4- Professor of Earth and Life Institute, Université catholique de Louvain, Louvain-la-Neuve, Belgium
DOI: 10.22067/JSW.2021.71748.1076

resulted in the highest accuracy and minimum error. Even though the DTr method presented R^2 values (0.52-0.39) weaker than the RF in the validation dataset, in general, the results of its spatial predictions were similar to the RF model from the surface to the subsurface and more stable than the k-NN. Against RF and DTr, k-NN couldn't display acceptable performance in the prediction of CCE% at all standardized depths.

Conclusion: In general, it is necessary to understand the spatial distribution of CCE due to its effect on soil moisture accessibility and plant nutrient uptake. Therefore, in the present study, we tried to introduce the RF machine learning algorithm as a superior model with environmental variables that were selected by PCA and the expert knowledge variable selection method. The maps prepared by this approach have an acceptable level of reliability for agricultural and environmental management by managers, soil experts, and farmers.

Keywords: Digital soil mapping, Standard depth, Spline function, Soil forming factors